Calculating spatial urban sprawl indices using open data

Luciano Gervasoni, Martí Bosch, Serge Fenet, Peter Sturm

Calculating spatial urban sprawl indices using open data

Luciano Gervasoni, Martí Bosch, Serge Fenet, Peter Sturm

Abstract

Urban sprawl has been related to numerous negative environmental and socioeconomic impacts. Meanwhile, urban areas have been growing at alarming rates, urging for assessing sprawl towards sustainable development. However, sprawl is an elusive term and different approaches to measure it have lead to heterogeneous results. Moreover, most studies rely on private/commercial data-sets and their software is rarely made public, impeding research reproducibility and comparability. Furthermore, many works give as result a unique value for a region of analysis, dismissing local spatial diversity that is vital for urban planners and policy makers. We present in this paper an open source framework for assessing urban sprawl using open data. Locations of residential and activity units are used to measure mixed use development and built-up dispersion, whereas the street network is used to measure accessibility between different land uses. Sprawl patterns are identified, and the resulting spatial information allows focusing on particular neighborhoods for a fine-grained analysis, as well as visualizing each sprawl dimension separately.

L. Gervasoni (Corresponding author) • P. Sturm
Inria, Univ. Grenoble Alpes, LJK, F-38000, Grenoble, France
Email: luciano.gervasoni@inria.fr, peter.sturm@inria.fr

M. Bosch
Ecole Polytechnique Fédérale de Lausanne, CEAT, Lausanne, Switzerland
Email: marti.bosch@epfl.ch

S. Fenet
Université de Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, F-69621, France
Email: serge.fenet@liris.cnrs.fr

1 Introduction

Urban sprawl is an elusive and ambiguous concept. “It is fair to say that the phenomenon is not fully understood to satisfaction in the academic, policy, or planning communities” (Torrens 2008).

As the number of people living in cities has been increasing considerably since 1950[1] (UN 2014), more than 66% of the world’s population are projected to live in urban areas by 2050, against 30% in 1950. The continuing population growth and urbanization are thus projected to add 2.5 billion people to the world’s urban population by 2050. In consequence, we face an increasing need for conceiving cities that can exist in a sustainable way.

Sprawl is an important phenomenon nowadays. “In Switzerland and Baden-Württemberg, Germany, at least as much land area has been taken up for settlement and transport within the 50 years between 1950 and 2000 as during the preceding 10,000 years” (Jaeger, Bertiller, Schwick, and Kienast 2010). “Milan has lost approximately 600,000 residents to the urban fringes over the last 15 years” (Hamidi and R. Ewing 2014). “From 2000-2006, Europe lost 1,117.9 km^2/y of natural and semi-natural areas to urban and other artificial land development” (Hennig et al. 2015). These facts urge to address the urban sprawl phenomenon in a pragmatical way, rather than solely focusing on a subjective discussion of its definition.

In order to cope with and reduce sprawl development in cities, several urban planning programs and concepts have been emerging. Among the most popular ones, one can mention Smart Growth, New Urbanism, Transit-Oriented Development and Traditional Neighborhood Development.

Yet, urban sprawl remains an elusive concept for two main reasons. First, its definition is subjective (Galster et al. 2001; Bruegmann 2005) and yields to no general consensus. Different authors formalize it differently, and sometimes even related works have found contradictory results when measuring the sprawl of given cities (Torrens 2008). However, a certain consensus exists on qualifying extreme cases of sprawl. For instance, a low level of sprawl is generally associated with compact cities with good street accessibility and featuring mixed use development. Second, one of the biggest constraints related to measuring urban sprawl is data availability, as the aspects of sprawl one can measure will be determined by the possibility of acquiring relevant data. Indeed, capturing all pertinent aspects of urban sprawl given limited data availability is still a challenge today.

[1] From 746 million to 3.9 billion in 2014.

Measuring sprawl is a key step towards the implementation of policies focusing on its limitations and the mitigation of its known negative effects. In this work we focus on the aspects of sprawl that impede sustainable development, and we present an open tool to assess the urban sprawl phenomenon using open data. The main founding principle of this work is to give city planners the availability to compute urban sprawl indices for any region in the world, therefore increasing the comparability of sprawl measurements. The measures of sprawl presented in this work are inspired by several related works reviewed in the following section. We focus on three recurring aspects of sprawl: dispersion, accessibility and land use mix.

In this work we provide neither a way to compute an integral measure of urban sprawl, nor a method to merge different dimensions of sprawl into one. Ideed, integral measures (a single value for an entire region of study) induce a loss of spatial information. We thus avoid reducing spatial indices into global metrics, which are always difficult to interpret. As for merging multiple measured dimensions into a single indicator, there exists no general consensus on how to do this in a plausible and objective manner. Further, we believe that this might also induce a loss of valuable information, hiding important underlying differences behind a single figure.

Given that there exists no general methodology for reducing dimensionality or performing spatial aggregation without losing valuable information, we argue that carrying out such procedures is up to the user, depending on the target application. For instance, if a study related to car dependency is conducted, it is possible to retrieve those aspects related to this dependency, while omitting others that are irrelevant to the study of interest.

Organization. In Sect. 2, a linkage between sprawl and sustainable development is recapitulated, followed by a description of related works on sprawl measurement. Sect. 3 details the input data we use, followed by the urban sprawl measurements in Sect. 4. Sect. 5 presents results, followed by conclusions.

2 Urban sprawl effects and sustainable development: related works

Urban sprawl has been linked to many negative effects in terms of sustainable development. Although some of them are still subject to discussion, most are now well known and we only cite them quickly. The reader can refer to the corresponding references.

- Water quality (Tu et al. 2007).
- Environmental costs and global warming effects (Johnson 2001; Gonzalez 2005; R. H. Ewing 2008).
- Air quality and pollution (Borrego et al. 2006; Stone 2008; Bhatta 2010).
- Inflated costs of infrastructure and public services (Carruthers and Ulfarsson 2003; R. H. Ewing 2008; Bhatta 2010; Hortas-Rico and Solé-Ollé 2010).
- Health effects (R. Ewing, Meakins, et al. 2014).
- Car dependency and traffic congestion (Newman, Kenworthy, and Vintila 1995; Cervero 1996; R. H. Ewing 2008).
- Loss of agricultural land and open space (R. H. Ewing 2008; Bhatta 2010).

It must be noted that the attribution of such effects to urban sprawl have been disputed in several works (remarkably Bruegmann 2005; Neuman 2005)

Several suitability criteria on sprawl indicators have been presented in (Jaeger, Bertiller, Schwick, and Kienast 2010). We find the first three criteria, the most important:

- Intuitive interpretation
- Mathematical simplicity
- Modest data requirements

In our case, we focus on developing practical sprawl measures which satisfies these criteria. An additional criterion might be added, namely that indicators shall actually be relevant to the measurement of sprawl...

Two basic questions arise when assessing urban sprawl:

- Geographical scope of analysis. The extreme cases are a stand-alone city case study, or world-wide coverage.
- Integral versus divisional measures, a categorization inspired by (Song, Merlin, and Rodriguez 2013). The extreme cases are a single metric for the whole region of study, or quasi-continuous spatial measures.

A comparison of the two extremes of geographical scope of analysis is provided in Tab. 1. The desired scope of analysis, as well as the local data availability will define the data-sets that can be used. Later, these data-sets will impose constraints on the specificity and/or the aspects of sprawl one can measure. In consequence, when assessing sprawl one should keep in

Case study	**General case**
− Stand-alone city	++ World-wide coverage
++ Data (precision, completeness, time-series availability)	−− Limited data (the higher the geographical scope, the stronger the limitations)
− Validation and pertinence of results (do theoretical assumptions hold?)	+ Evaluation and validation using heterogeneous cities. Potential for more pertinent conclusions
−− Comparability highly reduced	++ Comparable to other measures (related to geographical coverage)
−− Absolute metrics are imperative	++ Relative or absolute metrics (relative metrics appropriate if comparing enough cities)

Table 1: Comparison of the two extreme cases viz. geographical scope of sprawl analysis.

mind the constant trade-offs imposed by these scopes. For instance, one may focus on the analysis of a certain metropolitan area, for which much more detailed data on different aspects may exist, that would of course enrich the sprawl assessment. However, there is almost no possibility for comparing such a study with other studies of cities around the globe, unless the exact same data is available. The extracted sprawl features are specific to the type of data used for this area, disabling any valid comparison to other cities without introducing a possibly significant bias. For this type of studies, conclusions are generally only pertinent for cities within the scope of analysis.

On the other hand, data with world-wide coverage tend to be limited. The aspects of the complex phenomenon of sprawl that may be assessed using such data, are then constrained. Nevertheless, the usage of homogeneous data allows for worldwide comparisons on certain -albeit more limited- aspects of urban sprawl. Then, trends of certain widespread characteristics can be revealed and comparability is much easier, allowing for pertinent conclusions to be established.

Since the first work considering a multi-dimensional definition of urban sprawl (Galster et al. 2001), a number of works measuring and considering the multi-dimensional aspects of sprawl have emerged. A series of works have been developed and applied, mainly on cities in the United States (R. Ewing, Pendall, and Chen 2003; R. Ewing and Hamidi 2014; Hamidi and R. Ewing 2014; Hamidi, R. Ewing, et al. 2015). In these works, urban sprawl is mainly explained through four dimensions: Development density, land use mix, activity centering and street accessibility. This corresponds to a manageable number of dimensions in order to interpret the different aspects of sprawl. In the following, we review works considering the definition and measurement of these dimensions and discuss some of their shortcomings.

Several works provide sprawl indicators which rely on the assumption of the existence of a single Central Business District (CBD). This assumption is however not valid for every type of city. For example, polycentric cities naturally contain several CBD's, generally of smaller size compared to a monocentric city. Thus, low activity centering does not necessarily indicate sprawl. For instance, cities with low activity centering but with a high mix of uses may be considered to be little sprawled. The assumption on activity centering will hold in cities with clearly defined CBD's -such as several North American cities-, where the residential uses located elsewhere contain a low degree of mixed uses.

Indicators on street accessibility, when used to assess sprawl, should be related to traveling time or distance from households to activities. Thus, ad hoc indicators such as block size, number of 4-or-more intersections (R. H. Ewing 2008) do not seem immediately relevant to the assessment of sprawl. For example, the average graph-based road length between households and activities -of which the number of 4-or-more intersections is a weak proxy- might be more representative than block sizes to define the condition of accessibility related to urban sprawl. While a correlation may exist between such indicators and sprawl in cities of the United States, this might not be the case for cities elsewhere in the world, for example in Europe.

In this series of works, integral metrics are provided for a series of metropolitan areas. Spatial information is lost by delivering a single number for a whole metropolitan area that is difficult to interpret.

Comprehensive measures of sprawl are presented in (Torrens 2008). In this work, sprawl is divided in a series of characteristics or dimensions. For each of these dimensions, numerous attributes or aspects are measured. The study is carried out for cities in the United States. In addition, time-series analysis is done for a 10-year period. In this sense, the study is capable of capturing a great quantity of sprawl aspects. However, data requirements are not modest and for this reason, comparability is mainly limited the United States. Sprawl indices are presented in a spatially continuous visual way, greatly aiding to the interpretation of the captured sprawl.

Many works have focused on measuring urban sprawl by using satellite imagery as input. Remote-sensing based methods are capable of delivering urban form indices (Huang, Lu, and Sellers 2007; Sarvestani, Ibrahim, and Kanaroglou 2011; Poelmans and Van Rompaey 2009; Martellozzo and

Clarke 2011). But even though it is an important characteristic of sprawl, urban form is not the only dimension or aspect. This type of approaches also need to be complemented with analyses related to the effect of the street network and land usages.

Weighted Urban Proliferation (WUP) has been defined and used to assess sprawl in a series of works (Jaeger, Bertiller, Schwick, Cavens, et al. 2010; Jaeger and Schwick 2014; Hennig et al. 2015). Their studies are mainly centered on Europe and Canada. Here, sprawl assessment is based on three well-defined dimensions, namely urban permeation, dispersion and utilization density. However, during the composition of the final metric, an ad hoc weighting function is introduced for two of the three dimensions. While aiming at helping the visualization in extreme cases, this artificial weighting function has the contradictory effect of leading to a single final value that results from *a-posteriori* biased values. This goes in conflict with the first criterion of suitable measures for sprawl (Jaeger, Bertiller, Schwick, and Kienast 2010): intuitive interpretation.

Integral metrics are calculated at the NUTS-2 level for Europe, leading to a $1km^2$ grid level representation. Even though this is a low level of details, the necessary employment data demanded by utilization density rarely exist at such scale (except in this particular case for Switzerland). Thus, unless these data are available, the metric presented as theoretical utilization density, transforms, in practice, in a residential usage density.

3 Input data

We use OpenStreetMap[2] (OSM) data to calculate our urban sprawl indices. The main advantages of the OpenStreetMap dataset are:

- World-wide coverage.
- Homogeneity of data.
- Open data.
- Geo-localized data.
- Crowd-sourced data: potentially good precision and completeness (depending however on the region), rapid updates.
- Contains street network infrastructure, buildings and land use information.

[2] http://www.openstreetmap.org/

World-wide coverage, homogeneity and openness of data are very important aspects to help develop open and reproducible urban sprawl indices. Open data implies that these databases are freely available to everyone to use without restrictions by copyright, patents or other mechanisms of control. This will encourage comparisons and increase research reproducibility, leading to the clarification of a concept known to be elusive.

Crowd-sourced data are often more precise and complete than commercial data-sets and are updated more frequently. For instance, it has been concluded that the quality is “fairly accurate” for England (Haklay 2010), and it is even shown that OSM data are superior to the official data-set for Great Britain Meridian 2. This analysis has also been performed on Franch data (Girres and Touya 2010). Studies focusing on the street network of Germany have also been conducted in (Neis, Zielstra, and Zipf 2011), where it is concluded that the data-sets can be considered complete in relative comparison to a commercial data-set. In addition, the OSM data-set for Hamburg already covers about 99.8% of the street network (Over et al. 2010) according to the surveying office of Hamburg. In China, the volume of points of interest has been increasing substantially, e.g. nine-fold in the period 2007-2013 (Liu and Long 2016).

Furthermore, geo-localized data provide an advantage in comparison to gridded data, where spatial location specificity allows for finer-grained analysis.

Last but not least, quality metrics related to OSM data have been proposed in (Forghani and Delavar 2014; Barron, Neis, and Zipf 2014; Mooney, Corcoran, and Winstanley 2010; Fan et al. 2014), followed by different quality assessments, in particular for different countries. These metrics allow to assess beforehand the quality of the data to be used for a given sprawl analysis.

The main limitation or disadvantage related to OpenStreetMap is missing data. Though it highly varies for different cities in the world, there exists a relation between data completeness and city size. This can be explained by the fact that big cities hold, in general, a larger proportion of active contributors.

In the presented framework, we use data output from the *OSM2PGSQL* tool[3] that converts OpenStreetMap data to postGIS-enabled PostgreSQL databases. Shapefile formats allow for an easier representation and data man-

[3] `https://github.com/openstreetmap/osm2pgsql`

agement compared to the native node, way and relation OSM inner representation. This data is acquired from Mapzen Metro Extracts[4].

4 Sprawl Indices

We consider three main dimensions of sprawl: dispersion, land use mix, and accessibility, each leading to the computation of a dedicated numerical index. These indices are computed on a regular grid with an arbitrary, user-defined resolution, thus allowing for the computation of any point in any region of interest. The following sections describe the different sprawling dimensions used in this work.

4.1 Dispersion

We can define a sprawl-related urban dispersion in many ways. Taking into account our focus on sustainable development, we assume dispersion to be related to the density of urban form. It can be seen as the opposite of urban compactness. Our formal definition is inspired by the dispersion of built-up area defined in (Jaeger and Schwick 2014): “A landscape suffers from urban sprawl if it is permeated by urban development or solitary buildings [...]. The more area built over and the more dispersed the built-up area, [...] the higher the degree of urban sprawl.”

Effects. Too dispersed or scattered urban environments have been linked to an increase of costs of infrastructure and public services, loss of agricultural farmlands, and a trend to have an inefficient or nonexistent public transportation.

Rationale. The objective is to quantify the dispersion or scatteredness of the built environment, independently of actual land use. The intuition is to capture unused and lost spaces between buildings, which could, for example, have been better organized for leisure and agricultural uses (livability and sustainable uses). We thus ask the following question: “How dispersed is the location of buildings, losing valuable space in between which could have been employed for other uses?” This scatteredness will be directly linked to the relation between the space needed to set the built environment and the total amount of built-up area. In our metric, however, we voluntarily avoid

[4] `https://mapzen.com/metro-extracts/`

explicitly evaluating the built-up area itself given that it would penalize parks or green spaces around cities, which in fact are attractive and are one of the reasons that can refrain people from shifting to the suburbs. In consequence, we operationalize a proxy similar to the proximity index used in landscape metrics (McGarigal and Marks 1995). This closest-distance-based measure as follows.

Implementation. All polygons pertaining to buildings are retrieved from OpenStreetMap. For each building, the distance to its closest neighboring building is computed following the procedure depicted in 4.1.1. Then, for each node on the grid, a radius of search is considered and the median of the closest distance values for a local neighborhood (radius of search) is computed. The median was preferred to the mean due to the negative impact of outliers -e.g. a few isolated buildings-, which do not represent the dispersion of the majority of the buildings within the region of analysis, but that heavily impact the final value. The search radius is a parameter of the framework, experimentally set to $750m$ in the presented results.

4.1.1 Closest distance

For each polygon representing a building, the distance to the nearest neighboring building is computed, using the following methodology. First, latitude and longitude coordinates are projected to the Universal Transverse Mercator coordinate system (UTM). This system of coordinates provides a constant distance relationship anywhere on the map for a given zone number. For consistency, a unique zone number is used to project the coordinates. The center point of the encompassing bounding box of all latitude and longitude extracted coordinates is used to define the zone number for the UTM projection. A two-dimentional space-partitioning data structure (KD-tree) is created using the polygon's centroids, to speed up computations. Then, for each polygon P_r, the K nearest polygons are queried. The KD-tree data structure allows for very efficient axis-parallel range queries. Note that this will provide the K nearest centroids. Even though this does not necessarily mean they correspond to the exact nearest polygons, this first plausible approximation provide an efficient acceleration, and is anyhow corrected in the following step. Next, for each of the K neighbors, the real distance is evaluated using the actual associated polygons (the smallest distance between any two points on the polygon boundaries). Finally, the minimum is set as the closest distance to the reference polygon P_r. The parameter K is experimentally set to 50, ensuring that, even in the worst case, no polygon whose higher dimension is 50 times bigger than the dimensions of the neigbouring polygons will induce a bias.

4.2 Accessibility

Accessibility, in terms of sprawl, relates to the connectivity to activity uses (shops, amenities, leisure, health, commercial and industrial types, among others) given an input location. These input locations are usually associated to residential uses, e.g. households. Intuitively, accessibility answers the following question: "Do I need to travel long distances, in average, to reach different activity opportunities?"

Effects. Poor accessibility is related to car dependency, fossil fuel consumption, traffic jams, fine particles pollution, obesity and sedentary lifestyle.

Mostly, traveling behavior is related to the access to different activity uses. Car dependency of a certain household is directly related to the distance to these activities, whenever public transportation is ignored. Accessibility is related then, to the driving-journey which has to be done to reach different nearby activities.

Rationale. A location-based accessibility measure is operationalized using the street network infrastructure and all activity uses extracted from OpenStreetMap. We define a cumulative opportunities based measure: given an input location (potential household), and considering the street network, the number of activities that can be reached by traveling a pre-defined distance are counted. The higher the number of activities that can be accessed within this distance, the higher the degree of accessibility.

Cumulative opportunities based measures are popular in urban planning and geographical studies (Geurs and Van Wee 2004). As depicted in that study, easy interpretation and relatively low data requirements are the two major advantages related to this metric. These two aspects are important in our context, in order to explain the complex concept of sprawl and given the limited data availability, where no information related to travel behavior or infrastructure level of congestion is available.

Potential accessibility (gravity) measures are also widely used. However, they were not considered in this context given the increased complexity and difficulty of interpretation they impose (Geurs and Van Wee 2004).

We focus on the cumulative number of opportunities found within a travel distance X for two reasons:

- Short distances to the closest activities are captured by land use mixity (see below), where the co-occurrence of residential and activity uses are measured within a reduced local radius, meant for walkable distances.
- Longer distances -mostly related to car usage- are traveled in order to reach farther away, but still to reach necessary activity uses.

Implementation. First, the road infrastructure is extracted from OpenStreetMap and converted to a graph[5] Boeing 2017. Then, new nodes are created in this graph by splitting edges longer than a maximum tolerated length, set to $150m$ (see below for an explanation). Afterwards, each activity use extracted from OSM is associated to the closest node in the road network graph. The edge splitting allows to keep the approximation error of distances between activities and the road network, within acceptable limits.

Then, the accessibility is evaluated for each point of the grid. The closest graph node to that point is retrieved. If the distance between the grid point and the closest node is larger than a given threshold, the accessibility is not measured. In the results presented here, this parameter threshold is set to $250m$. This threshold is used to explicitly ignore the particular cases for which an artificially low accessibility should be computed (around parks, forests, or mountainous areas, for example).

The distance X is experimentally set to $1km$, even though it can be modified and adapted to other values according to the particular traveling behavior in each city. The underlying idea is to define a desired distance value for traveling on a daily basis. The higher the number of activities associated within the distance X, the smaller the average traveling distances. This is explained by the assumption that the more activity opportunities lie nearby, the lower should be the need to travel long distances on a daily basis.

4.3 Land use mix

Mixed use development is a type of urban development that blends residential and activity (commercial, industrial, amenities, shops, institutional) uses. By land use mix, we denote the close-by co-occurrence of such different land uses.

[5] `https://github.com/gboeing/osmnx`

Effects. A poor land use mix has been shown to favor several negative outcomes. Among the most important ones are inefficient transportation, car dependency, and health issues such as obesity and sedentary lifestyle (Song, Merlin, and Rodriguez 2013).

Rationale. The calculation of a land use mix indicator, based on the Entropy Index, is adopted from our previous work, see (Gervasoni et al. 2016) for details. The goal of this measure is to determine to what extent the spatial configuration of land uses is well distributed in a city, assessing the co-occurrence of different land uses in local neighborhoods.

5 Results

The software implementation of the present work is openly available[6]. It allows the easy comparison of a set of cities across the world. Results are depicted here using the UTM coordinates.

Given the current lack of ground truth data for assessing urban sprawl, there exists no methodology for validating sprawling indices. Consequently, validation is currently done via manual inspection. In this paper, we compute indices for both sprawled and compact cities. The cities we use as case studies here are Grenoble (France), Manhattan (New York), and Mays Chapel (Maryland). Grenoble is a European city known to have a non-sprawling city center, with sprawling villages around its mountain-bordered surroundings. Manhattan is known to be compact, with a very good accessibility and mixed use development. It represents therefore the inverse of a sprawling city, whereas Mays Chapel is a typical sprawling city. This selection allows us to test distinct attributes of sprawl, as well as sprawl structures at different scales.

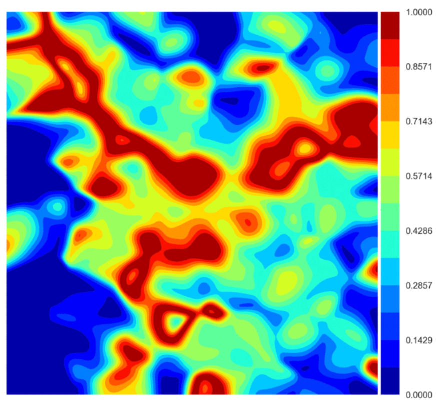
All of the indices are calculated on regular grids with a resolution of $100m$. For the accessibility indices, a multi-processing version has been implemented due to its computational load. Plus, the number of processed nodes were 29083, 33293, and 13794 for respectively Manhattan, Grenoble and Mays Chapel.

Grenoble. Fig. 1 shows the computed sprawl indices. The results reveal a good land use mixture around the city center, and along the city major axes. This mixture is reduced drastically in the mountainous surroundings (North,

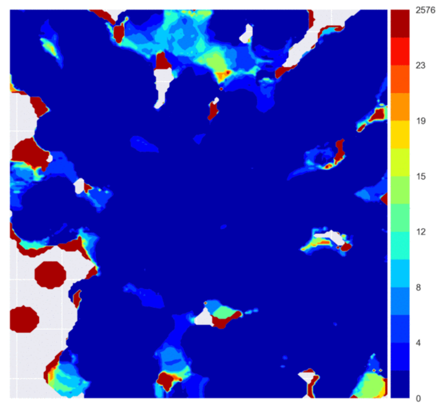
[6] `https://gitlab.inria.fr/gervason/urbansprawl`

West and Southeast of the city center, cf. the areas with few roads in Fig. 1 (d)). The city is found to be mainly compact. High dispersion values are seen only in the mountainous regions.

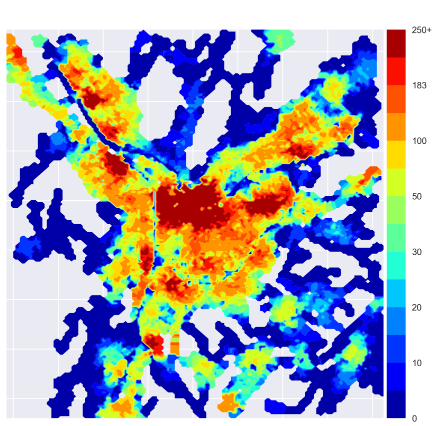
Finally, high accessibility values are found around the city. These values are reduced drastically when driving around its mountainous roads. Note that the spatial indices allow to visualize the degree of sprawl, found to increase from the city center towards mountainous locations.



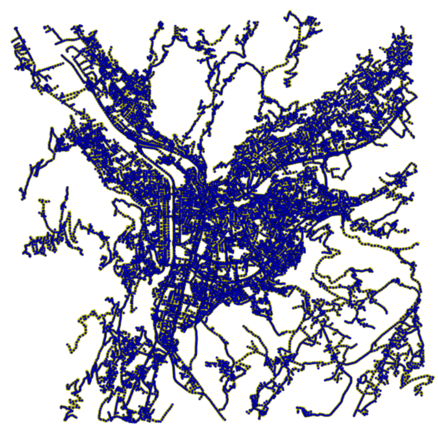
(a) Land use mix indices



(b) Dispersion indices



(c) Accessibility indices



(d) Street network

Fig. 1: Grenoble sprawl indices.

Manhattan. Fig. 2 depicts the computed sprawl indices. In accordance with what might be expected for an extremely low-sprawled city, Manhattan shows a high degree of mixed use development and accessibility, while dispersion is very low. Even the Brooklyn part of the city (to the Southeast of Manhattan island) features very good accessibility and compactness values, and a relatively high land use mix.

On should remark that our method faces its limitations when depicting land use mix for cities with very tall buildings, as in Midtown Manhattan. Our method currently does not include the intensity of a land use in the computation of land use mix: residential buildings are counted the same way regardless of their height or number of apartments. This effect can be witnessed in Fig. 3 a and b, that represent the local density of activities and residential housings respectively. The residential density of center Manhattan if artificially low. Relevant data (e.g. number of floors) are filled more and more frequently in the OpenStreetMap database[7], hence we will take into account land use intensity in future work.

Mays Chapel. See Fig. 4 for computed sprawl indices. Mays Chapel is located North of Baltimore and features a typical residential-sprawling pattern.

Accessibility values are generally low in Mays Chapel's city core and higher along the highway axis, where an increased location of activity uses are found. The built environment at some distance to the highway axis is found to be increasingly dispersed. This results in a typical pattern of the "American Dream" residential development.

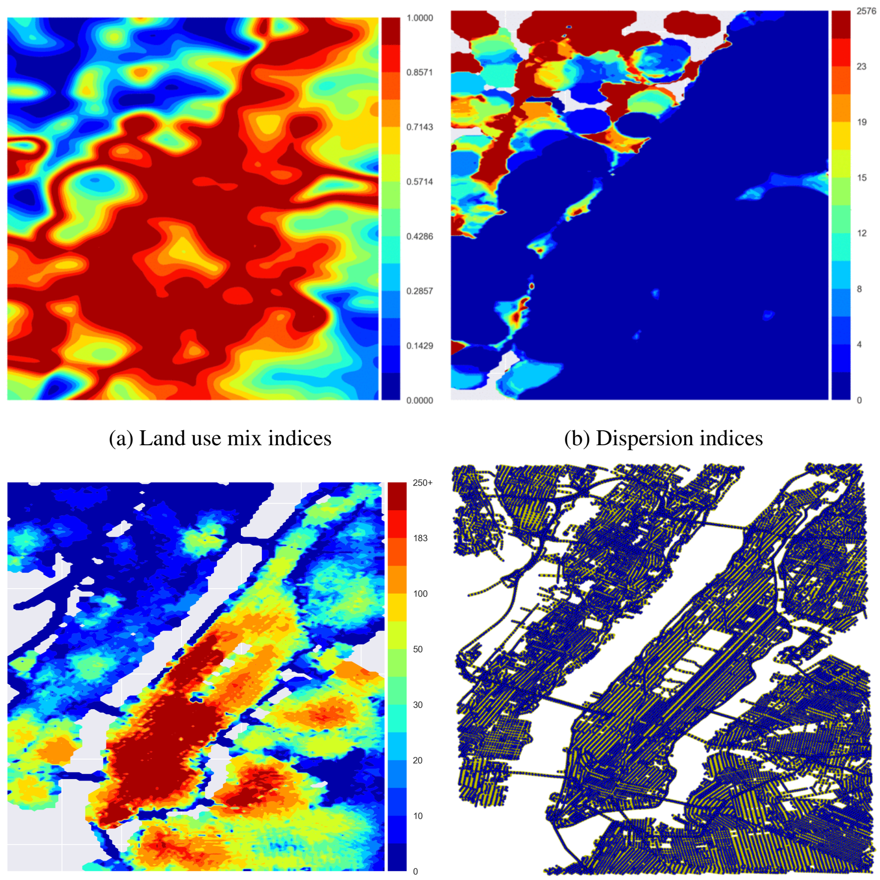
Brooklandville is located Southwest of Mays Chapel. Many offices, shops, restaurants, banks and indoor sport facilities are found there. The results clearly capture this village's characteristics with high accessibility and land use mix, as expected.

Table 2 shows the time needed to calculate each procedure of the sprawling indices. Tests were carried out using an Intel Xeon E5-2609 2.40 GHz with 4 cores under a typical operating system load.

6 Conclusion

An open framework that assess the urban sprawl phenomenon in the context of sustainable development has been presented. The three considered

[7] See `http://wiki.openstreetmap.org/wiki/Buildings`



(a) Land use mix indices

(b) Dispersion indices

(c) Accessibility indices

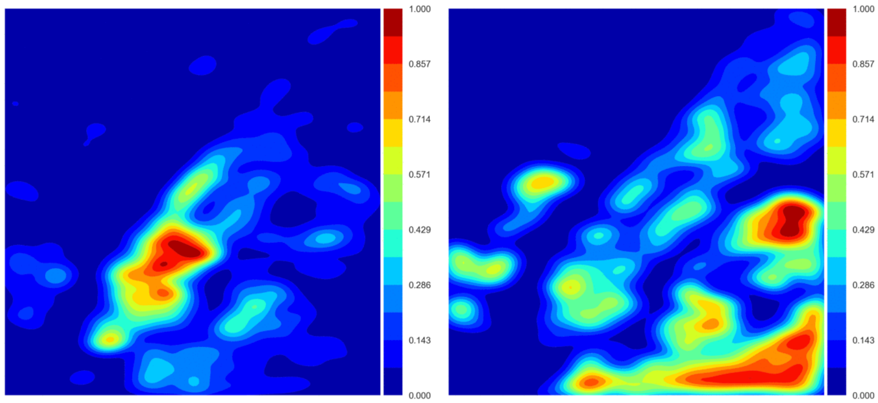
(d) Street network

Fig. 2: Manhattan sprawl indices.

dimensions of sprawl permit a manageable quantity of aspects to analyze.

Process / Time (minutes)	**Grenoble**	**Manhattan**	**Mays Chapel**
OSM data extraction & classification	9	14	1
Land use mix	7	28	1
Dispersion	11	8	4
Accessibility	139	144	27

Table 2: Computational time (minutes) for sprawl indices calculation.



(a) Activity uses densities

(b) Residential uses densities

Fig. 3: Manhattan land usage densities.

In addition, each of them is defined in a simple, easy to interpret and as pertinent as possible way.

A comparison between cities in different continents is depicted, a novelty in the literature on urban sprawl indices. This might encourage further comparison between cities across the world.

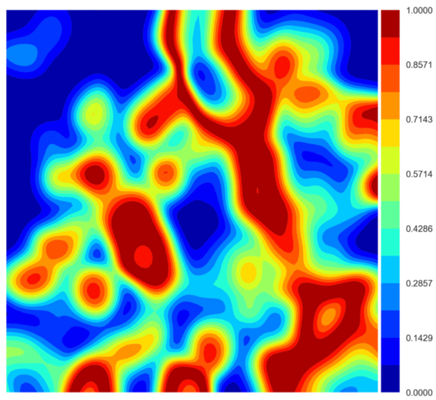
Divisional sprawling indices are provided at a fine-grained analysis for an input city. These spatial indices admit an intra-city analysis for assessing the degree of sprawl in local neighborhoods. This type of analysis is not possible in the case of integral measures.

Nonetheless, the presented work still has limitations. OpenStreetMap data are still far from complete for many cities across the world, although this is rapidly improving. Also, as explained above, the land use mix computation does not yet take into account the intensity of land uses. First steps to remedy this will be to consider building heights to weight residential and activity uses, as well as to include fine-grained worldwide gridded population data[8].

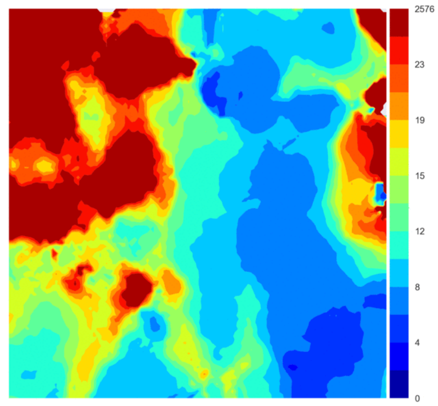
Acknowledgements

We would like to thank the computing centre CNRS/IN2P3 (Lyon-Villeurbanne), for supplying the computing resources necessary for the presented tool.

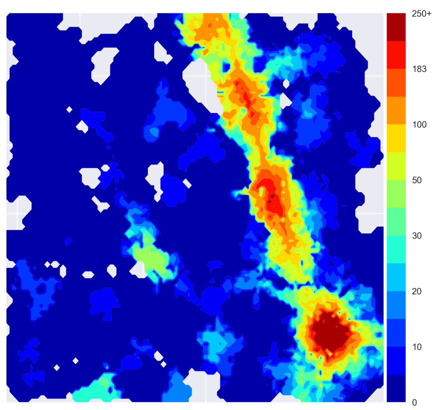
[8] `http://sedac.ciesin.columbia.edu/data/collection/gpw-v4`



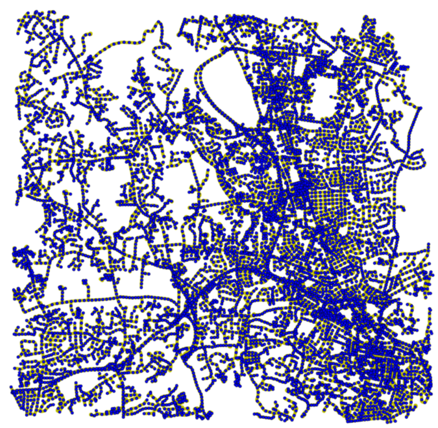
(a) Land use mix indices



(b) Dispersion indices



(c) Accessibility indices



(d) Street network

Fig. 4: Mays Chapel, Maryland, sprawl indices.

References

Barron, Christopher, Pascal Neis, and Alexander Zipf (2014). "A comprehensive framework for intrinsic OpenStreetMap quality analysis". In: *Transactions in GIS* 18.6, pp. 877–895.

Bhatta, Basudeb (2010). "Causes and consequences of urban growth and sprawl". In: *Analysis of Urban Growth and Sprawl from Remote Sensing Data*. Springer, pp. 17–36.

Boeing, G. (2017). *OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks*. Under review. DOI:

10.2139/ssrn.2865501. URL: http://geoffboeing.com/publications/osmnx-complex-street-networks/.

Borrego, Carlos et al. (2006). "How urban structure can affect city sustainability from an air quality perspective". In: *Environmental Modelling & Software* 21.4, pp. 461–467.

Bruegmann, Robert (2005). *Sprawl: a compact history*. The University of Chicago Press.

Carruthers, John I and Gudmundur F Ulfarsson (2003). "Urban sprawl and the cost of public services". In: *Environment and Planning B: Planning and Design* 30.4, pp. 503–522.

Cervero, Robert (1996). "Mixed land-uses and commuting: evidence from the American Housing Survey". In: *Transportation Research Part A: Policy and Practice* 30.5, pp. 361–377.

Ewing, Reid H (2008). "Characteristics, causes, and effects of sprawl: A literature review". In: *Urban Ecology*. Springer, pp. 519–535.

Ewing, Reid and Shima Hamidi (2014). *Measuring urban sprawl and validating sprawl measures*. Tech. rep. National Institutes of Health, Ford Foundation, Smart Growth America.

Ewing, Reid, Gail Meakins, et al. (2014). "Relationship between urban sprawl and physical activity, obesity, and morbidity–update and refinement". In: *Health & Place* 26, pp. 118–126.

Ewing, Reid, Rolf Pendall, and Don Chen (2003). "Measuring sprawl and its transportation impacts". In: *Transportation Research Record: Journal of the Transportation Research Board* 1831, pp. 175–183.

Fan, Hongchao et al. (2014). "Quality assessment for building footprints data on OpenStreetMap". In: *International Journal of Geographical Information Science* 28.4, pp. 700–719.

Forghani, Mohammad and Mahmoud Reza Delavar (2014). "A quality study of the OpenStreetMap dataset for Tehran". In: *ISPRS International Journal of Geo-Information* 3.2, pp. 750–763.

Galster, George et al. (2001). "Wrestling sprawl to the ground: defining and measuring an elusive concept". In: *Housing Policy Debate* 12.4, pp. 681–717.

Gervasoni, Luciano et al. (2016). "A framework for evaluating urban land use mix from crowd-sourcing data". In: *2nd International Workshop on Big Data for Sustainable Development*. URL: http://hal.inria.fr/hal-01396792.

Geurs, Karst T and Bert Van Wee (2004). “Accessibility evaluation of land-use and transport strategies: review and research directions”. In: *Journal of Transport Geography* 12.2, pp. 127–140.

Girres, Jean-François and Guillaume Touya (2010). “Quality assessment of the French OpenStreetMap dataset”. In: *Transactions in GIS* 14.4, pp. 435–459.

Gonzalez, George A (2005). “Urban sprawl, global warming and the limits of ecological modernisation”. In: *Environmental Politics* 14.3, pp. 344–362.

Haklay, Mordechai (2010). “How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets”. In: *Environment and Planning B: Planning and Design* 37.4, pp. 682–703.

Hamidi, Shima and Reid Ewing (2014). “A longitudinal study of changes in urban sprawl between 2000 and 2010 in the United States”. In: *Landscape and Urban Planning* 128, pp. 72–82.

Hamidi, Shima, Reid Ewing, et al. (2015). “Measuring sprawl and its impacts: An update”. In: *Journal of Planning Education and Research* 35.1, pp. 35–50.

Hennig, Ernest I et al. (2015). “Multi-scale analysis of urban sprawl in Europe: Towards a European de-sprawling strategy”. In: *Land Use Policy* 49, pp. 483–498.

Hortas-Rico, Miriam and Albert Solé-Ollé (2010). “Does urban sprawl increase the costs of providing local public services? Evidence from Spanish municipalities”. In: *Urban Studies* 47.7, pp. 1513–1540.

Huang, Jingnan, Xi X Lu, and Jefferey M Sellers (2007). “A global comparative analysis of urban form: Applying spatial metrics and remote sensing”. In: *Landscape and Urban Planning* 82.4, pp. 184–197.

Jaeger, Jochen AG, Rene Bertiller, Christian Schwick, Duncan Cavens, et al. (2010). “Urban permeation of landscapes and sprawl per capita: New measures of urban sprawl”. In: *Ecological Indicators* 10.2, pp. 427–441.

Jaeger, Jochen AG, Rene Bertiller, Christian Schwick, and Felix Kienast (2010). “Suitability criteria for measures of urban sprawl”. In: *Ecological Indicators* 10.2, pp. 397–406.

Jaeger, Jochen AG and Christian Schwick (2014). “Improving the measurement of urban sprawl: Weighted Urban Proliferation (WUP) and its application to Switzerland”. In: *Ecological Indicators* 38, pp. 294–308.

Johnson, Michael P (2001). “Environmental impacts of urban sprawl: a survey of the literature and proposed research agenda”. In: *Environment and Planning A* 33.4, pp. 717–735.

Liu, Xingjian and Ying Long (2016). “Automated identification and characterization of parcels with OpenStreetMap and points of interest”. In: *Environment and Planning B: Planning and Design* 43.2, pp. 341–360.

Martellozzo, Federico and Keith C Clarke (2011). “Measuring urban sprawl, coalescence, and dispersal: a case study of Pordenone, Italy”. In: *Environment and Planning B: Planning and Design* 38.6, pp. 1085–1104.

McGarigal, Kevin and Barbara J Marks (1995). *FRAGSTATS: spatial pattern analysis program for quantifying landscape structure*. Tech. rep. USDA Forest Service. URL: `http://www.umass.edu/landeco/research/fragstats/fragstats.html`.

Mooney, Peter, Padraig Corcoran, and Adam C Winstanley (2010). “Towards quality metrics for OpenStreetMap”. In: *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, pp. 514–517.

Neis, Pascal, Dennis Zielstra, and Alexander Zipf (2011). “The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011”. In: *Future Internet* 4.1, pp. 1–21.

Neuman, Michael (2005). “The compact city fallacy”. In: *Journal of Planning Education and Research* 25.1, pp. 11–26.

Newman, Peter, Jeffrey Kenworthy, and Peter Vintila (1995). “Can we overcome automobile dependence?: Physical planning in an age of urban cynicism”. In: *Cities* 12.1, pp. 53–65.

Over, M et al. (2010). “Generating web-based 3D City Models from OpenStreetMap: The current situation in Germany”. In: *Computers, Environment and Urban Systems* 34.6, pp. 496–507.

Poelmans, Lien and Anton Van Rompaey (2009). “Detecting and modelling spatial patterns of urban sprawl in highly fragmented areas: A case study in the Flanders–Brussels region”. In: *Landscape and Urban Planning* 93.1, pp. 10–19.

Sarvestani, Mahdi Sabet, Ab Latif Ibrahim, and Pavlos Kanaroglou (2011). “Three decades of urban growth in the city of Shiraz, Iran: A remote sensing and geographic information systems application”. In: *Cities* 28.4, pp. 320–329.

Song, Yan, Louis Merlin, and Daniel Rodriguez (2013). “Comparing measures of urban land use mix”. In: *Computers, Environment and Urban Systems* 42, pp. 1–13.

Stone, Brian (2008). “Urban sprawl and air quality in large US cities”. In: *Journal of Environmental Management* 86.4, pp. 688–698.

Torrens, Paul M (2008). “A toolkit for measuring sprawl”. In: *Applied Spatial Analysis and Policy* 1.1, pp. 5–36.

Tu, Jun et al. (2007). “Impact of urban sprawl on water quality in eastern Massachusetts, USA”. In: *Environmental Management* 40.2, pp. 183–200.

UN (2014). *World Urbanization Prospects: The 2014 Revision, Highlights*. Tech. rep. United Nations, Population Division, Department of Economic and Social Affairs.